主题热度加速度指数*
——学科研究热点识别新方法

■ 荣国阳 李长玲 范晴晴 郭凤娇
山东理工大学信息管理研究院 淄博 255049

摘 要:[目的/意义]横向与纵向综合识别不同时间阶段的研究热点,有助于把握学科发展历程和方向,为热点主题识别方法拓展研究思路。[方法/过程]构建累积主题热度模型 TP,反映主题在学科中的横向相对研究热度;构建主题热度加速度指数模型 TAI,量化研究主题发展的纵向速度变化情况;构建学科研究热点识别模型 TP * TAI,从横向和纵向两方面综合反映主题研究热度及其变化情况。[结果/结论]采用图情领域 2001 - 2020 年的研究文献为样本进行实证分析,结果表明:该模型可以有效识别各时间阶段的研究热点,并将其区分为前沿、稳定和衰退 3 种类型,实现学科研究的动态描述。

关键词:主题热度 研究热点 识别方法 TP * TAI 模型
分类号:G250
DOI:10.13266/j.issn.0252 - 3116.2021.20.007

1 引言

一篇文献的关键词或主题词是其核心内容的浓缩和提炼,代表该文献的研究主题。如果某一关键词或主题词在其所在学科的文献中反复出现,则说明该关键词或主题词所表现的研究主题是该学科的研究热点[1],也是研究人员广泛关注和研究的高热度研究主题[2]。学科在不同时间阶段有不同的研究热点,这些热点代表了该阶段学科研究的重点和发展方向。识别学科热点及其变化情况,既有助于研究人员把握学科发展历程和方向、选择研究主题、合理分配研究资源,同时也为热点主题评价、未来发展趋势预测及相关规律发现等提供研究基础。目前,构建科学方法识别学科研究热点成为把握学科现状并预测未来趋势的关键[3-4],国内外诸多学者从不同角度选择不同方法识别学科研究热点。

首先,运用关键词突现、共现、引文分析方法,识别学科研究热点。K. Mane 和 K. Bo¨rner[5] 运用 Kleinberg 突发检测算法和共词分析方法,识别美国国家科学院院刊论文的基因突变、分子序列、蛋白质研究等热点主题;刘小慧等[6]改进 TF-IDF 算法,识别 2015 年情报学的研究热点为用户研究、大数据等;高继平等[7]、汤强等[8]、胡秀梅等[9]采用多词共现与聚类分析方法,识别数字信息传输、3D 打印技术、图书情报战略规划等不同学科领域的研究热点;Y. Chang 等[10]、X. Ping[11]、D. Rossetto 等[12]、C. Jebari 等[13]采用引文耦合和共引分析方法,识别图书情报学、国际抗癌研究、商业管理、生物医学等领域的研究热点。

其次,运用时间序列分析法或时间加权方法,识别学科研究热点及其变化趋势。肖婷婷等[14]采用时间序列可视化分析方法,研究发现语义标注的研究以本体和知识库为基础,随着语义网研究的深入而不断发展;刘自强等[15]提出基于时间序列模型的研究热点评价和预测方法,分析 2005 - 2014 年竞争情报研究现状与发展趋势,预测 2015 年研究目标并进行验证,发现模型是可行有效的;周鑫等[16]构建词频变化率模型,分析 2000 - 2014 年 CSSCI 收录的情报学领域期刊文献,识别增长、稳定和下降型研究热点;李长玲等[17]基于时间因子改进 z 指数,对比 z 指数改进前后主题热度的排名变化,将 2014 - 2018 年情报学研究热点划分

* 本文系国家社会科学基金重点项目“跨学科潜在知识生长点识别与创新趋势预测研究”(项目编号:19ATQ006)研究成果之一。
作者简介:荣国阳(ORCID:0000 - 0002 - 5822 - 2306),硕士研究生;李长玲(ORCID:0000 - 0001 - 6266 - 4820),教授,硕士,硕士生导师,通讯作者,E-mail:lichl69@163.com;范晴晴(ORCID:0000 - 0003 - 3593 - 0470),硕士研究生;郭凤娇(ORCID:0000 - 0002 - 2902 - 8299),馆员,博士。
收稿日期:2021 - 03 - 23 修回日期:2021 - 06 - 04 本文起止页码:59 - 67 本文责任编辑:杜杏叶

为上升、稳定和下降 3 种类型；奉国和等[18]构建时间加权关键词词频分析模型，识别 CSSCI 中图书情报领域期刊论文的研究热点，并将结果区分为上浮、下降和稳定型关键词；J. Li 等[19]基于时间视角的共词网络提出聚类系数和路径长度计算方法，识别 SCI、SSCI、CPCI-S 和 CPCI-SSH 四个数据库中生物能源领域的研究热点。

最后，采用计算机算法和模型识别学科研究热点。孙海生[20]基于超网络模型改进共词和共引分析方法，识别 2014－2016 年图书情报学研究热点，包括大数据类、Altmetrics 类、移动图书馆类等 7 类；C. Figuerola 等[21]、X. Han[22]应用主题建模统计技术和隐狄利克雷分配模型，识别不同学科的热点主题；阮光册等[23]采用 Doc2Vec 方法对论文摘要进行向量计算，并分析其相似度，生成热点选题论文集，提取主题描述并识别教育学热点主题，包括高等教育改革、高等教育公平、学习方式等；裘惠麟等[24]综合 LDA2vec 模型和 Word2Vec 词向量化、文档向量化模型提出多源数据的热点识别方法，识别机器学习领域的热点主题包括文本分类和特征检测等。

综上所述，目前研究热点的识别主要采用词频突现、共现聚类、引文分析、词频变化率、时间加权、模型算法等方法，从某一视角识别学科热点研究主题，但未对主题热度的变化情况进行深入分析。因此，本文在分析学科研究主题横向热度的同时，分析热度加速度的纵向变化情况，综合识别学科不同阶段的研究热点。步骤如下：①构建累积主题热度模型 TP，即学科内某研究主题累积词频占比，反映某时间段内研究主题的横向相对研究热度；②构建主题热度加速度指数模型 TAI，表示主题热度增速变化情况，反映研究主题在时间视角下的纵向热度及其变化；③构建学科研究热点识别模型 TP * TAI，综合反映学科内研究主题横向相对研究热度与纵向变化情况，识别各时间段研究热点；④制定分类标准，将热点主题细分为前沿、稳定和衰退 3 种类型，以准确把握学科发展动态。

2 基于主题热度加速度的学科热点识别模型构建

本文构建 TP * TAI 模型，横纵向综合识别学科热点主题。其中，累积主题热度模型 TP 反映学科内的主题横向热度，主题热度加速度指数模型 TAI 反映纵向热度变化情况。

2.1 累积主题热度模型 TP

累积主题热度（Topic Popularity，TP）模型用某时间段某研究主题累积词频（即研究文献量）在学科文献总量中的占比表示，表达式为：

$$TP=\frac{\sum_{t=n}^{i}C_t}{\sum_{t=n}^{i}P_t}\ (n\leqslant i\leqslant m) \qquad 公式(1)$$

式中，t 为年份，C_t 为 t 年某研究主题的研究文献量，P_t 为 t 年的学科文献总量。n 为该研究主题第一次出现的年份或数据分组的起始年份，m 为最近的年份或数据分组的截止年份。

TP 模型采用相对累积量对研究主题热度进行逐年测度，不仅可以反映主题发展至 t 年的热度情况，还可以消除各年份文献绝对数量不同导致的误差。虽然该模型可以反映某时间段某研究主题在该学科的横向相对研究热度，但累积计算的方法无法表达主题热度的变化趋势。因此，本文构建主题热度加速度指数模型 TAI，以发现主题研究热度的纵向变化趋势。

2.2 主题热度加速度指数模型 TAI

在物理学概念中，加速度是指速度变化量与发生这一变化所用时间的比值。其表达式为：

$$a=\frac{v_2-v_1}{\Delta t} \qquad 公式(2)$$

其中，v_2-v_1 指速度变化量，Δt 指两次速度变化的间隔时间。若加速度 a 大于零，则表明物体沿正方向做加速运动；若加速度 a 小于零，则物体在正方向做减速运动；若加速度 a 等于零，则物体静止或做匀速运动。

公式（1）中，若 m = n，表示某主题 t 年的相对热度，即：

$$TP_t=\frac{C_t}{P_t} \qquad 公式(3)$$

本文将加速度概念引入研究热点识别中，用于测度主题研究热度的加速变化情况，那么 TP_t 可在相对累积热度中看作 t 年的相对增长速度。那么，时间间隔一年（$\Delta t=1$）的主题热度加速度 a 可以表示为：

$$a=\frac{TP_t-TP_{t-1}}{\Delta t}=\frac{C_t}{P_t}-\frac{C_{t-1}}{P_{t-1}} \qquad 公式(4)$$

与物理中加速度的表达含义相同，若 a > 0，说明该主题研究热度呈加速增长趋势；若 a = 0，说明该主题热度增长速度不变；若 a < 0，说明该主题热度增长但增长速度减缓。显然，不论主题热度加速度 a 取何值，主题研究热度始终不小于零。为表达主题热度加速度与主题热度的关系，本文对主题热度加速度 a 取

以 e 为底的指数，得到主题热度加速度指数（Topic Acceleration Index，TAI）模型，表达式为：

$$TAI = e^{a} = e^{\frac{C_t}{P_t} - \frac{C_{t-1}}{P_{t-1}}} \quad \text{公式(5)}$$

式中，TAI 指数始终大于零，符合主题研究热度始终不小于零的特点。若 a＞0，则 TAI＞1；若 a＜0，则 0＜TAI＜1；若 a＝0，则 TAI＝1。

由于 TAI 模型测度主题热度的加速度，而非速度。所以，该模型一方面能够识别累积文献量不高而加速度极高的研究主题，这类主题的研究热度短时间内急剧增长，有可能是新的知识生长点；另一方面还可以识别累积文献量较高而加速度为负的研究主题，这类主题的研究热度短时间内急速下降，筛除研究热点之列。因此，该模型的优势在于，能够感知研究主题短时间内的热度变化，有效把握学科发展趋势。

2.3 学科热点识别模型 TP＊TAI

累积主题热度 TP 考量了某主题相较于其他研究主题在学科中的横向相对研究热度，主题热度加速度指数 TAI 反映了某主题自身在学科发展过程中的纵向热度变化情况。本文构建学科研究热点识别模型如下：

$$TP * TAI = \frac{\sum_{t=n}^{i} C_t}{\sum_{t=n}^{i} P_t} \times e^{\frac{C_t}{P_t} - \frac{C_{t-1}}{P_{t-1}}} (n \leqslant i \leqslant m) \quad \text{公式(6)}$$

该模型中，累积主题热度 TP 反映某时间段内某研究主题在学科领域中的横向相对热度，主题热度加速度指数 TAI 反映主题热度增长速度的变化情况，TP＊TAI 将二者结合，从横向与纵向两方面综合评价主题的研究热度。因此，TP＊TAI 模型能够反映主题的累积研究热度和动态变化情况，从宏观上识别时间跨度较长的学科稳定型核心热点主题，从微观上识别学科领域的阶段性新兴热点主题和当前前沿热点主题。

3 实证分析——以图书馆学情报学为例

3.1 数据来源与预处理

本文以图书馆学情报学为例，以 CNKI 期刊数据库为数据源，样本来源期刊选择依据为：2019 年综合影响因子排名前十的 CSSCI 期刊。由于 CNKI 中《情报学报》2003－2012 年数据缺失，无法作为样本，最终选择的样本期刊有《中国图书馆学报》《图书情报知识》《大学图书馆学报》《图书与情报》《情报理论与实践》《图书情报工作》《情报资料工作》《情报科学》《情报杂志》9 种期刊。采集 9 种期刊 2000－2020 年 55 444 篇载文的题录信息，将 2000 年的数据作为 TAI 指数计算的第一组 TP_{t-1}，2001－2020 年的数据用作学科研究热点识别数据。数据采集时间为 2021 年 1 月 12 日。去除 55 444 篇载文中的卷首语、编辑部公告、征稿启事等非研究论文 2 127 篇，得到有效样本 53 317 篇。

本文选择论文作者标识的关键词为该文的研究主题，用以检验模型的有效性、识别案例学科的研究热点。对样本文献数据进行预处理：①用 bibexcel 对关键词字段进行拆分；②统计每年的关键词及其词频；③筛选有效词。去除研究目的不明确的词，如“影响因素”“分析”等；去除表示研究背景的词，如“美国”“中国”等；合并同义近义关键词，如合并“新冠肺炎”“新型冠状病毒肺炎”等为“新冠肺炎”。

3.2 学科热点识别

按以下步骤对预处理后的关键词样本数据进行处理：

（1）构建年份－关键词词频矩阵。取每年词频阈值大于等于 5 的关键词共 1 359 个，将关键词词频与年份的对应数据 C_t，以及每年的发文总量 P_t，汇总至表 1（部分结果）。

（2）累积主题热度 TP 计算。为描述我国图书情报学学科发展细节，同时避免计算结果失真，将 2001－2020 年的数据以 5 年为一组进行划分，共 4 组。将表 1 中数据分别代入公式（1），并令 n＝分组起始时间，m＝分组截止时间，计算 TP 值，部分计算结果见表 2。

（3）主题热度加速度指数 TAI 计算。由于主题热度每年的相对增长速度 TP_t 数值都较小，$TP_t - TP_{t-1}$ 没有较好的区分度，导致主题热度加速度指数 TAI 接近于 1。为较好区分主题研究热度的影响，达到反映主题增长速度变化的目的，对 TP_t 放大千倍处理。为方便计算，将公式（5）调整为 $TAI = e^{\frac{C_t}{P_t \times 10} - \frac{C_{t-1}}{P_{t-1} \times 10}}$，以实现 TP_t 放大处理。将表 1 中数据分别代入该公式，*TAI* 部分计算结果见表 3。

（4）学科热点识别及分组排序。将表 2 数据与表 3 数据分别对应相乘，得到每个主题词在 2001－2020 年间每年的 *TP＊TAI* 值，共 20 个。将这 20 个值按每 5 年分为一组，共 4 组。取组内平均值作为某主题在该时间段的综合得分。依据综合得分对全部主题词进行组内排序，部分计算结果见表 4。

表 1 年份 – 关键词词频矩阵(部分)

指标类型	关键词	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
C_t	图书馆	159	241	277	183	267	219	253	233	187	218	268	266	202	179	135	113	88	74	68	50	44
	信息服务	75	131	123	88	113	92	71	73	74	74	80	55	58	47	31	32	24	19	18	11	7
	Internet	74	75	59	33	6	5	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0
	高校图书馆	73	139	126	97	81	90	84	93	92	106	143	122	103	115	105	120	114	99	70	69	66
	数字图书馆	53	149	220	180	177	170	117	107	88	90	84	116	95	52	49	41	24	19	25	16	9
	知识管理	27	40	76	65	100	122	123	122	106	109	111	89	54	44	25	29	19	18	13	8	6
	知识经济	57	47	37	20	12	9	6	4	2	3	1	0	0	0	0	1	1	0	0	1	0
	信息资源	50	80	76	56	80	79	70	43	63	44	64	28	29	31	12	5	4	9	8	4	1
	网络环境	43	83	86	78	66	32	37	18	22	19	15	13	5	8	2	4	2	1	2	1	0
	因特网	40	39	27	18	20	10	10	2	0	1	0	0	0	0	0	0	0	0	0	0	0
	数据库	34	53	48	37	42	23	25	16	15	15	17	13	15	10	13	5	2	2	1	0	4
	资源共享	33	31	20	15	30	20	34	23	28	12	25	14	11	17	11	4	7	4	6	3	3
	图书馆学	31	41	42	40	45	43	44	51	33	43	33	45	27	17	31	22	8	13	12	16	10
	信息	29	29	26	30	31	26	26	12	10	13	8	15	9	4	5	2	2	3	7	2	1
	竞争情报	27	24	33	30	39	51	65	73	60	90	90	55	75	47	51	42	30	21	20	16	10
	信息产业	26	33	20	8	18	4	12	8	9	9	5	2	2	2	1	1	4	1	0	0	0
	信息技术	25	32	47	20	22	24	19	22	15	15	16	23	13	12	3	6	3	6	3	1	2
	信息检索	25	40	33	33	71	46	48	39	43	34	29	25	33	13	24	13	14	7	6	5	2
	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
	大数据	0	0	0	0	0	0	0	0	0	0	0	0	8	35	55	77	106	85	68	72	65
	新冠肺炎	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	83
	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
P_t	当年总文献量	1 794	2 414	2 559	2 369	2 700	2 584	2 702	2 522	2 496	3 123	3 316	3 352	2 883	2 556	2 430	2 291	2 195	2 055	1 952	1 894	1 979

表 2 累积主题热度 TP 计算结果(部分)

关键词	2001 – 2005 年					2006 – 2010 年					2011 – 2015 年					2016 – 2020 年				
	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
图书馆	0.100	0.104	0.095	0.096	0.094	0.094	0.093	0.087	0.082	0.082	0.079	0.075	0.074	0.070	0.066	0.040	0.038	0.037	0.035	0.032
信息服务	0.054	0.051	0.047	0.045	0.043	0.026	0.028	0.028	0.027	0.026	0.016	0.018	0.018	0.017	0.017	0.011	0.010	0.010	0.009	0.008
Internet	0.031	0.027	0.023	0.017	0.014	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
高校图书馆	0.058	0.053	0.049	0.044	0.042	0.031	0.034	0.035	0.035	0.037	0.036	0.036	0.039	0.040	0.042	0.052	0.050	0.046	0.043	0.041
数字图书馆	0.062	0.074	0.075	0.072	0.071	0.043	0.043	0.040	0.037	0.034	0.035	0.034	0.030	0.028	0.026	0.011	0.010	0.011	0.010	0.009
知识管理	0.017	0.023	0.025	0.028	0.032	0.046	0.047	0.046	0.042	0.040	0.027	0.023	0.021	0.019	0.018	0.009	0.009	0.008	0.007	0.006
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
大数据	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.005	0.009	0.013	0.048	0.045	0.042	0.041	0.039
新冠肺炎	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.041
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

表 3 主题热度加速度指数 TAI 计算结果(部分)

关键词	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
图书馆	7.3E+04	4.5E+03	3.4E−14	2.5E+10	7.2E−07	7.2E+03	0.287	2.6E−08	0.006	6.1E+04	0.231	9.2E−05	0.966	5.2E−07	0.002	9.8E−05	0.017	0.309	2.1E−04	0.015
信息服务	2.5E+05	2.0E−03	1.8E−05	1.1E+03	0.002	8.9E−05	14.418	2.018	0.002	1.537	4.4E−04	40.846	0.177	0.004	3.355	0.048	0.185	0.976	0.032	0.103
Internet	3.0E−05	3.3E−04	1.0E−04	8.2E−06	0.750	0.209	1.026	0.672	1	1.352	0.996	0.742	1	1	1	1	1	1	1	1
高校图书馆	2.1E+07	2.4E−04	2.5E−04	1.8E−05	125.176	0.023	326.17	0.983	0.054	9.7E+03	0.001	0.512	1.1E+04	0.168	9.6E+03	0.642	0.023	4.5E−06	1.768	0.046
数字图书馆	9.5E+13	3.4E+12	4.6E−05	2.9E−05	1.263	2.0E−10	0.417	4.7E−04	0.001	0.031	1.1E+04	0.191	3.8E−06	0.835	0.103	9.5E−04	0.185	35.221	0.012	0.020
知识管理	4.571	5.0E+05	0.104	1.5E+04	2.6E+04	0.184	17.331	0.027	5.1E−04	0.239	0.001	4.0E−04	0.220	0.001	10.699	0.018	1.108	0.123	0.087	0.303
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
大数据	1	1	1	1	1	1	1	1	1	1	1	16.037	5.5E+04	7.6E+03	5.8E+04	2E+06	0.001	0.001	24.016	0.005
新冠肺炎	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.6E+18
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

注:* 表中 7.3E+04 即 7.3×10^4

表4　4组TP＊TAI综合得分排序及学科研究热点识别结果(部分)

排名	2001－2005年		2006－2010年		2011－2015年		2016－2020年	
	关键词	综合得分	关键词	综合得分	关键词	综合得分	关键词	综合得分
1	数字图书馆	1.16E＋12	图书馆	1.13E＋03	大数据	218.741	新冠肺炎	6.81E＋16
2	图书馆	4.83E＋07	Web2.0	104.599	高校图书馆	161.986	大数据	2.29E＋04
3	高校图书馆	2.49E＋05	高校图书馆	73.379	网络舆情	106.934	突发公共卫生事件	3.90E＋03
4	信息服务	2.80E＋03	本体	54.255	数字图书馆	73.804	人工智能	81.998
5	知识管理	2.60E＋03	信息资源	17.179	微博	65.581	情报工作	44.641
6	信息检索	827.670	公共图书馆	10.473	竞争情报	62.023	科学数据	13.690
7	网络环境	229.149	企业	5.195	社会网络分析	7.563	扎根理论	13.521
8	情报学	21.320	图书馆学教育	2.504	情报学	4.113	网络舆情	7.610
9	电子商务	18.933	竞争情报	1.991	引文分析	1.647	情报学	1.912
10	知识产权	8.426	知识共享	1.697	知识服务	1.211	情报感知	1.827
11	WTO	7.607	指标体系	1.689	图书馆学	1.046	文献计量	1.568
12	信息资源	3.628	知识产权	1.503	科学数据	0.685	数字人文	1.275
13	企业	2.384	情报学	1.446	专利分析	0.655	数据治理	1.233
14	元数据	2.053	知识服务	1.370	学科馆员	0.638	智库	1.157
15	合并	1.612	大学图书馆	0.927	云计算	0.625	阅读推广	1.109
16	文献检索	0.761	开放存取	0.591	可视化	0.577	图书情报学	0.527
17	竞争情报	0.618	信息管理	0.587	期刊评价	0.511	图书馆学教育	0.514
18	电子政务	0.567	知识转移	0.535	移动图书馆	0.458	共词分析	0.513
19	信息技术	0.537	资源共享	0.469	关联数据	0.423	内容分析	0.503
20	数据挖掘	0.519	引文分析	0.407	信息检索	0.336	开放数据	0.425
21	信息化	0.401	读者服务	0.322	共词分析	0.305	区块链	0.307
22	搜索引擎	0.326	图书馆2.0	0.313	聚类分析	0.294	应急管理	0.264
23	引文分析	0.286	图书馆学	0.252	阅读推广	0.273	信息生态	0.208
24	文献计量学	0.280	数字资源	0.194	公共图书馆	0.206	网络谣言	0.199
25	参考咨询	0.277	危机管理	0.191	开放获取	0.186	政府数据	0.196
26	信息需求	0.252	图书馆联盟	0.179	知识图谱	0.185	服务质量	0.172
27	网络资源	0.239	图书馆服务	0.171	信息服务	0.175	社会网络分析	0.170
28	资源共享	0.228	知识管理	0.170	学位论文	0.160	研究热点	0.167
29	enterprise	0.224	信息公平	0.166	学科服务	0.155	信息行为	0.164
30	电子阅览室	0.205	政府信息公开	0.158	文献计量学	0.121	情感分析	0.139

3.3　学科研究热点分类

本文根据各研究主题排名变化情况，将研究热点分为前沿型、稳定型和衰退型3类。罗瑞等[2]通过概念辨析及特性研究，认为研究前沿具有近期产生和高创新价值两个特征；郑彦宁等[25]认为前沿是相对于特定研究领域、特定时间而言的，可以代表研究领域最新的研究进展或动向；颜端武等[26]从主题演化的角度，将研究前沿定义为新出现的、有发展潜力的研究主题；Q. Wang[27]认为新兴前沿主题具有新颖性、快速增长、连贯性、高影响力和不确定性5个特点。

综合上述观点，本文认为前沿研究热点应当具备近期出现或突显、增长迅速、高影响力三个特点。因此，研究热点3种类型的界定如下：①前沿研究热点。排名位于最近5年(2016－2020年)中，样本研究主题总量的前2%，且在最近两组数据中突显或首次出现，表达为2016－2020年或2010－2015年较前一组排名上升50%以上；②稳定研究热点。学科研究中综合得分持续靠前的主题，表达为4组综合得分均位于样本研究主题总量的前10%，且相邻两组数据的排名波动不超过样本研究主题总量的5%；③衰退研究热点。曾在某时间段成为学科发展的热点主题，但随后热度持续降低的主题，即4组综合得分排名呈递减趋势，且极差大于样本研究主题总量的20%。

设某研究主题在第k组(1≤k≤4)的排名为Rk，样本研究主题总量为j。那么，上述分类依据可表达为图1。

编写VBA程序，按图1所示分类原则，将表4中学科热点识别结果区分为3类，分别取前8位，结果如

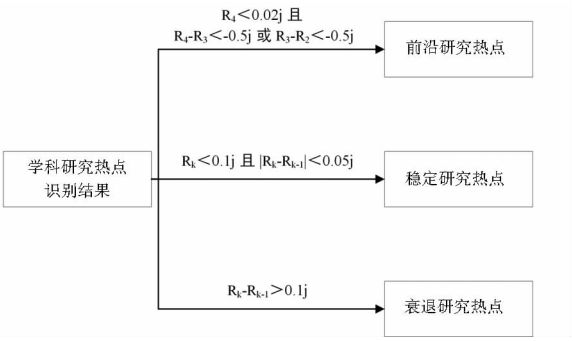

图 1 研究热点分类原则

表 5 所示：

表 5 图书情报学前沿、稳定和衰退研究热点

类型	研究热点
前沿型	新冠肺炎、大数据、突发公共卫生事件、人工智能、情报工作、科学数据、扎根理论、网络舆情
稳定型	情报学、高校图书馆、引文分析、本体、公共图书馆、电子政务、数据挖掘、文献计量
衰退型	信息服务、知识管理、信息检索、电子商务、知识产权、信息资源、竞争情报、信息技术和信息化

“前沿型”研究热点是图书情报学最近几年的高热度研究主题，具有学科研究热度占比高、增速快的特点，有较大研究潜力和发展势头；“稳定型”研究热点具有宏观上热度相对稳定、微观上波动发展的特点，是情报学研究较为稳固的核心研究内容；“衰退型”研究热点具有热度逐年下降的趋势，表示该类主题在发展过程中积累了一定的研究成果，已相对成熟，近年该类研究或更加深入细致，或转向相近领域。

4 模型效果分析

从表 1－4 分析发现：模型 TP＊TAI 可以从横向与纵向两方面有效识别学科热点，TP 测度研究主题某时间段的相对热度，TAI 反映某主题研究热度随时间的变化情况；对比时间段相近的相关前期研究结果发现：TP＊TAI 模型不仅能通过年度累积数据分析长时间跨度的学科热点宏观情况，还能识别不同时间段的研究热点，包括前沿、稳定和衰退型热点主题。

4.1 模型有效性

（1）累积主题热度模型 TP 可以有效反映主题在某时间的相对热度。例如表 1 第 1、2 行，词频矩阵中“图书馆”在 2017 年、“信息服务”在 2008 和 2009 年的绝对词频相同，均为 74；然而，受学科研究文献总量不同的影响，TP 值所反映的累积主题热度计算结果明显不同，表 2 第 1、2 行中“图书馆”在 2017 年的累积主题热度 TP 为 0.038，而“信息服务”在 2008 年与 2009 年的累积主题热度 TP 分别为 0.028 与 0.027。可见，该模型消除了因各年文献总量不同而产生的测度误差，可以反映主题发展至测度年份的累积热度情况。

（2）主题热度加速度指数模型 TAI 可以有效反映某主题的研究热度随时间变化情况。表 3 第 6 行，“知识管理”在 2002、2004、2005 年的 TAI 值远大于 1，数级达到万级，代表这 3 年主题热度的加速度极大，说明该时间段“知识管理”的研究热度激增；但从 2008 年开始，除 2015 年和 2017 年外，该研究主题的 TAI 值均小于 1，代表其研究热度加速度为负，说明该时间段“知识管理”的增速下降，进入衰退期。可见，该模型引入的加速度维度可以分时段有效识别主题热度的增速情况，弥补 TP 模型无法体现主题热度纵向发展变化的缺陷，可以有效体现主题研究热度的加速变化情况，实现动态分析。

（3）TP＊TAI 模型综合累积主题热度 TP 和热度加速度 TAI 两个指标，从横向与纵向两个角度综合测度主题热度及其变化情况，有效识别各时期的研究热点。例如，表 2 中 2011－2015 年“大数据”的 TP 值明显小于“数字图书馆”，说明这个时间段“大数据”在情报学中的研究热度低于“数字图书馆”；表 3 中 2012－2015 年“大数据”的热度加速度 TAI 远大于“数字图书馆”，说明“大数据”的研究态势、热度趋势高于“数字图书馆”；综合横向与纵向的表现，表 4 中 2011－2015 年“大数据”TP＊TAI 综合排名高于“数字图书馆”，分列第 1 和第 4 位。另外，“新冠肺炎”作为 2020 年首次出现的高热度研究主题，表 2、表 3 中 TP 和 TAI 值都很大，代表该主题研究热度高、热度增长速度也高，表 4 中 TP＊TAI 计算结果高达 6.81×1016，是该年度最热门的研究主题。因此，TP＊TAI 模型中两个指标能够有效影响综合排名，且各时间阶段排名靠前的主题词均有两个特点：主题研究文献量在学科总文献量中占比较高，即累积主题热度较高；呈高速增长趋势，即主题热度加速度较大。

4.2 模型优势

本文将 2001－2020 年时间跨度的图书情报学研究文献，以 5 年为一组划分为 4 个时间段，从宏观与微观综合识别学科热点。为了更进一步分析模型效果，发现其优势，选择每个时间段相近的研究文献，对比分析研究结论。由于目前还未有 2016－2020 年时间段图书情报学研究热点的相关文献，因此选择 2001－2005、2006－2010、2011－2015 三个时间段相近的同类研究识别结果进行对比分析，相关数据见表 6。

表 6　同类研究识别结果对比

同类研究	识别时间段	同类研究识别结果	TP＊TAI 筛除	TP＊TAI 补充
邱均平等[28]	1999－2007	图书馆信息服务、信息检索、数字图书馆、知识管理、文献计量、竞争情报	无	知识产权、WTO、电子政务等
王敬兰[29]	2004－2009	图书馆学情报学理论、学术评价、学科馆员与虚拟参考咨询、知识管理与知识服务、知识组织与信息检索、信息资源共建共享、图书馆事业与建设、数字图书馆信息资源、Web2.0	数字图书馆、信息服务等	本体、竞争情报等
王知津等[30]	2010－2014	情报学理论研究、文献计量研究、竞争情报研究、网络舆情研究、微博研究	无	大数据、云计算、科学数据等

以上学者通过词频或共现分析得出的不同阶段的研究结论，与本文表 4 中同时间段的研究结果相同或相近，说明本文 TP＊TAI 识别模型是可行有效的。同时，与前期研究相比，本文分时段细化研究粒度，从横向与纵向两方面综合反映主题的研究热度与趋势，更具优势。

（1）能分析微观、中观与宏观学科发展变化情况。本文将时间跨度细分，一方面能把握每个时间段的学科研究热点分布，另一方面能了解长时间跨度的学科发展脉络和历程。如：表 2、表 3 能分析学科年度热点及其变化；表 4 能识别不同时间段的研究热点，把握学科阶段性及其在样本长时间跨度内的发展变化情况。

（2）能识别到各时间段研究热度在学科中占比不高，但增速较快的新兴研究主题；也能识别并筛除学科占比不低，且增速缓慢的渐进衰退研究主题。

2001－2005 年，与同期文献[28]相比，本文除了涵盖前期研究成果的识别结果外，还识别到 WTO/知识产权/电子商务等当期热点研究主题。这些突显研究主题受科研环境影响，如：1998 年启动“首都电子商务工程”，2000 年中国电子商务协会在京成立[31]，2001 年中国加入 WTO，虽然当期研究文献总量学科占比不高，但是研究热度增速较快，所以能被本文很好地识别。

2006－2010 年，本文部分识别结果与同期文献[29]相同，如 web2.0/信息资源/图书馆学情报学学科研究等。也补充识别到当期不同的研究热点，如：2004 年，图书情报领域开始了关于“本体”的研究[32]，并在 2006－2010 年阶段研究热度迅速升温，所以本文有效识别到了“本体”这一新兴主题。同时，学科馆员/虚拟参考咨询等是图书馆传统服务研究领域，研究文献量占比不低，但研究热度增速较低，所以被本文识别并筛除。

2011－2015 年，与同期文献[30]相比，除涵盖前期研究成果的识别结果外，还识别到大数据等当期热点研究主题。2012 年 3 月，奥巴马政府发布《大数据研究和发展倡议》[33]，2015 年武汉大学召开“大数据时代图书情报学理论与教育发展对策”国际研讨会[34]，以信息数据管理为核心研究内容的图书情报学科，迅速引入大数据的研究，并热度猛增，使“大数据”的研究热度在该时期热度排名第一。

（3）能识别学科目前的前沿热点主题。2020 年新冠疫情席卷全球，2017 年国家社科重大项目“情报学学科建设与情报工作未来发展路径研究”立项，2018 年国务院办公厅印发实施《科学数据管理办法》[35]等。因此，本文识别新冠疫情/突发公共卫生事件/情报工作/科学数据等是 2016－2020 年热点主题，也是图书情报学科的前沿问题。大数据/人工智能等持续增长的研究热度，依旧是图书情报学的前沿热点问题。

综上，通过文献数据验证与专家咨询，发现本文识别的各时期研究热点，比较符合图书情报学学科发展的实际情况，说明模型是可行有效的。

5　结语

本文构建学科研究热点识别模型 TP＊TAI。其中，累积主题热度模型 TP 反映学科内横向相对研究热点，主题热度加速度指数模型 TAI 反映时间视角下的纵向研究热度变化情况。实证表明，该模型具有以下特征：①从横向与纵向两个角度反映各时间阶段研究主题的相对热度及其变化情况，有助于把握学科的中、长期发展历程与方向；②有效识别图书情报学 20 年来的“前沿型”“稳定型”和“衰退型”研究热点，实现学科研究的动态分析。

虽然模型在图书情报学领域得到有效验证，但在后续研究中，一方面要多维度对模型进行验证和完善，提高模型在其他学科应用的普适性；另一方面还要从标题、摘要、全文中抽取代表论文研究主题的关键词，使学科热点识别更全面、有效。

参考文献：

[1] 邱均平，温芳芳. 近五年来图书情报学研究热点与前沿的可视化分析——基于 13 种高影响力外文源刊的计量研究[J]. 中

国图书馆学报, 2011,37(2):51 - 60.
[2] 罗瑞, 许海云, 董坤. 领域前沿识别方法综述[J]. 图书情报工作, 2018,62(23):119 - 131.
[3] GLANZEL W, THIJS B. Using 'core documents' for detecting and labelling new emerging topics[J]. Scientometrics, 2012,91(2):399 - 416.
[4] TU Y, SENG J. Indices of novelty for emerging topic detection [J]. Information processing & management, 2012,48(2):303 - 325.
[5] MANE K, BRNER K. Mapping topics and topic bursts in PNAS [J]. Proc natl acad sci USA, 2004,101(S1):5287 - 5290.
[6] 刘小慧, 李长玲, 冯志刚. 基于改进的 TF * IDF 方法分析学科研究热点:以情报学为例[J]. 情报科学, 2017,35(7):82 - 87.
[7] 高继平, 丁堃, 潘云涛, 等. 多词共现分析方法的实现及其在研究热点识别中的应用[J]. 图书情报工作, 2014,58(24):80 - 85, 98.
[8] 汤强, 王亚民, 赵艳. 基于 g 指数和共现指数的研究热点及合作团体分析[J]. 情报杂志, 2014,33(9):72 - 75.
[9] 胡秀梅, 高凡. 国内图书情报领域图书馆战略规划研究热点探析[J]. 图书情报工作, 2016,60(9):13 - 17, 27.
[10] CHANG Y, HUANG M, LIN C. Evolution of research subjects in library and information science based on keyword, bibliographical coupling, and co-citation analyses[J]. Scientometrics, 2015,105(3):2071 - 2087.
[11] XIE P. Study of international anticancer research trends via co-word and document co-citation visualization analysis[J]. Scientometrics, 2015,105(1):611 - 622.
[12] ROSSETTO D, BERNARDES R, BORINI F, et al. Structure and evolution of innovation research in the last 60 years: review and future trends in the field of business through the citations and co-citations analysis[J]. Scientometrics, 2018(115):1329 - 1363.
[13] JEBARI C, VIEDMA E, COBO M. The use of citation context to detect the evolution of research topics: a large-scale analysis[J]. Scientometrics, 2021,126(4):2971 - 2989.
[14] 肖婷婷, 邱均平, 祖旋, 等. 语义标注研究热点与演进历程的知识图谱分析[J]. 情报理论与实践, 2015,38(1):1 - 6, 22.
[15] 刘自强, 王效岳, 白如江. 基于时间序列模型的研究热点分析预测方法研究[J]. 情报理论与实践, 2016,39(5):27 - 33.
[16] 周鑫, 陈媛媛. 关键词词频变化视角下学科研究发展趋势分析——以国内情报学研究为例[J]. 情报杂志, 2016,35(5):133 - 140.
[17] 李长玲, 牌艳欣, 相富钟, 等. 改进 z 指数的高被引学科研究热点识别方法探讨[J]. 情报理论与实践, 2020,43(6):69 - 75, 96.
[18] 奉国和, 孔泳欣. 基于时间加权关键词词频分析的学科热点研究[J]. 情报学报, 2020,39(1):100 - 110.
[19] LI J, WANG Y, YAN B. The hotspots of life cycle assessment for bioenergy: a review by social network analysis[J]. Science of the total environment, 2018,625:1301 - 1308.
[20] 孙海生. 基于超网络模型的研究热点探测与聚类主题描述[J]. 情报杂志, 2017,36(6):93 - 98.
[21] FIGUEROLA C G, MARCO F J G, PINTO M. Mapping the evolution of library and information science (1978 - 2014) using topic modeling on LISA[J]. Scientometrics, 2017,112: 1507 - 1535.
[22] HAN X. Evolution of research topics in LIS between 1996 and 2019: an analysis based on latent dirichlet allocation topic model [J]. Scientometrics, 2020(125):2561 - 2595.
[23] 阮光册, 夏磊. 基于 Doc2Vec 的期刊论文热点选题识别[J]. 情报理论与实践, 2019,42(4):107 - 111.
[24] 裘惠麟, 邵波. 多源数据环境下科研热点识别方法研究[J]. 图书情报工作, 2020,64(5):78 - 88.
[25] 郑彦宁, 许晓阳, 刘志辉. 基于关键词共现的研究前沿识别方法研究[J]. 图书情报工作, 2016,60(4):85 - 92.
[26] 颜端武, 苏琼, 张馨月. 基于时序主题关联演化的科学领域前沿探测研究[J]. 情报理论与实践, 2019,42(7):144 - 150.
[27] WANG Q. A bibliometric model for identifying emerging research topics[J]. Journal of the association for information science and technology, 2018,69(2):1 - 25.
[28] 邱均平, 周春雷, 杨思洛. 改革开放 30 年来我国情报学研究的回顾与展望(三)——情报学的发展阶段及趋势分析[J]. 图书情报研究, 2009,2(3):1 - 9.
[29] 王兰敬. 2004 - 2009 年我国图书馆、情报与档案管理学科的研究热点与重点领域——基于 CSSCI 来源文献关键词的分析[J]. 图书情报工作, 2011,55(16):68 - 71.
[30] 王知津, 李博雅. 我国情报学研究热点及问题分析——基于 2010 - 2014 年情报学核心期刊[J]. 情报理论与实践, 2016,39(9):7 - 13.
[31] 邱均平, 马秀娟. 1998 - 2009 年国内电子商务研究论文的计量分析[J]. 情报科学, 2011,29(5):641 - 646.
[32] 李健康, 张春辉. 本体研究及其应用进展[J]. 图书馆论坛, 2004(6):80 - 86.
[33] 杨绎. 基于文献计量的"大数据"研究[J]. 图书馆杂志, 2012,31(9):29 - 32.
[34] 郭晓婉, 冉从敬, 吴丹, 等. 大数据时代图书情报学理论与教育发展对策——第四届中美数字时代图书馆学情报学教育国际研讨会综述[J]. 图书情报知识, 2016(1):116 - 121.
[35] 国务院办公厅. 国务院办公厅关于印发科学数据管理办法的通知[EB/OL]. [2021 - 05 - 23]. http://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm.

作者贡献说明:

荣国阳:研究方法设计,数据处理,论文撰写与修改;
李长玲:论文框架设计,论文指导,论文撰写与修改;
范晴晴:数据采集,论文修改与校对;
郭凤娇:论文指导,论文修改与校对。

Topic Acceleration Index: A New Method for Identifying Discipline Research Hotspots

Rong Guoyang　Li Changling　Fan Qingqing　Guo Fengjiao

Institute of Information Management, Shandong University of Technology, Zibo 255049

Abstract: [**Purpose/significance**] The horizontal and longitudinal comprehensive identification of research hotspots at different time stages is helpful to grasp the development course and direction of the discipline, and expands research ideas for hot topic identification methods. [**Method/process**] The cumulative topic popularity model TP was constructed to reflect the horizontal research heat of the topic in the discipline. The topic acceleration index model TAI was built to quantify the longitudinal speed change of the development of the research topic. TP * TAI model was constructed to identify the hot spots of subject research, which reflected the research heat and its change comprehensively from both horizontal and longitudinal aspects. [**Result/conclusion**] The empirical analysis on the literatures of LIS from 2001 to 2020 shows that the model can effectively identify the research hotspots in different time stages, and distinguish three types, namely frontier, stable and declining, to realize the dynamic description of the subject research.

Keywords: topic popularity　research hotspots　identification method　TP * TAI model

《图书情报工作》杂志社发布出版伦理声明

为加强和增进学术论文写作、评审和编辑过程中的学术规范、科研诚信与学术道德建设，树立良好学风，弘扬科学精神，坚决抵制学术不端，建立和维护公平、公正、公开的学术交流生态环境，《图书情报工作》杂志社（包括《图书情报工作》《知识管理论坛》两个期刊编辑部）结合两刊实际，特制订出版伦理声明并于2020年2月正式发布。

该出版伦理声明承诺两刊将严格遵守并执行国家有关学术道德和编辑出版相关政策与法规，规范作者、同行评议专家、期刊编辑等在编辑出版全流程中的行为，并接受学术界和全社会的监督。共包括三大部分，总计十五条，分别为：一、作者的出版伦理（①学术论文是科学研究的重要组成部分；②学术不端是学术论文的毒瘤；③作者是学术论文的主要贡献者；④ 作者署名体现作者的知识产权与学术贡献；⑤学术论文要高度重视知识产权与信息安全；⑥参考文献的规范性引用是学术规范的重要表征；⑦要高度重视研究数据与管理的规范性；⑧建立纠错与学术自我净化机制）。二、同行评议专家的出版伦理（⑨同行评议是论文质量的重要控制机制；⑩评审专家应遵守论文评审的相关要求；⑪评审专家要严格遵循相关的伦理指南和行为准则）。三、编辑的出版伦理（⑫编辑应成为学术论文质量的守护者；⑬编辑应在学术道德建设中发挥监控作用；⑭编辑要成为遏制学术不端的最后屏障；⑮对学术不端实行“零容忍”）。

全文请见：http://www.lis.ac.cn/CN/column/column291.shtml

（本刊讯）